Genomics 99 (2012) 233–240

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative genomic analysis of two brucellaphages of distant origins

Victor Flores [a], Ahidé López-Merino [b], Guillermo Mendoza-Hernandez [c], Gabriel Guarneros [a,*]

[a] Departamento de Genética y Biología Molecular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV IPN), México Distrito Federal, Mexico
[b] Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional (ENCB IPN), México Distrito Federal, Mexico
[c] Departamento de Bioquímica, Facultad de Medicina de la Universidad Nacional Autónoma de México, México Distrito Federal, Mexico

ARTICLE INFO

Article history:
Received 22 November 2011
Accepted 3 January 2012
Available online 21 January 2012

Keywords:
Brucella
Comparative genomics
DnaA/*oriC* DNA replication
DNA packaging clade

ABSTRACT

Here, we present the first complete genome sequence of brucellaphage Tbilisi (Tb) and compared it with that of Pr, a broad host-range brucellaphage recently isolated in Mexico. The genomes consist of 41,148 bp (Tb) and 38,253 bp (Pr), they differ mainly in the region encoding structural proteins, in which the genome of Tb shows two major insertions. Both genomes share 99.87% nucleotide identity, a high percentage of identity among phages isolated at so globally distant locations and temporally different occasions. Sequence analysis revealed 57 conserved ORFs, three transcriptional terminators and four putative transcriptional promoters. The co-occurrence of an ORF encoding a putative DnaA-like protein and a putative *oriC*-like origin of replication was found in both brucellaphages genomes, a feature not described in any other phage genome. These elements suggest that DNA replication in brucellaphages differs from other phages, and might resemble that of bacterial chromosomes.

1. Introduction

The genus *Brucella* comprises Gram-negative rod-shaped bacteria capable of infecting mammal hosts, including man, cattle, goats and small ruminants. Human brucellosis is a serious public health concern, and animal brucellosis causes substantial economical losses [1]. The worldwide distribution and prevalence of brucellosis make it the most widespread zoonosis [2]. The genus consists of nine species, all of which are closely related and show very low levels of genetic diversity [3]. *Brucella* cells can be infected by lytic phages known as brucellaphages. These are short-tailed particles with dsDNA as genetic material. Most brucellaphages are host specific, a reason why a robust typing system was developed using several reference brucellaphages [4]. Brucellaphages are very similar to each other in morphology, antigen reactions and overall physicochemical properties [4–6]. In fact, restriction mapping and hybridization of their DNAs have failed to detect significant differences despite the geographical remoteness and temporal separation of the isolates [7,8]. Research on *Brucella* bacteriophages has been conducted for more than 60 years; however, little is known about their genome organization, evolution and distribution [8]. Here, we compared the complete genome sequences of two brucellaphages isolated more than 50 years apart in geographically remote locations. To our knowledge, this is the first time that the complete genome of brucellaphages is reported. We observed that in spite of the remote origin of the phages, the nucleotide sequence of their DNAs was well conserved. Among the differences between the genomes, there were two insertions in the genome of phage Tb and several single nucleotide polymorphisms.

2. Results and discussion

2.1. Plaque morphology and host range of brucellaphages Tb and Pr

Brucellaphage Tb was isolated in Tbilisi, Georgia, in 1955 from manure. It is a short-tailed bacteriophage consisting of double-stranded genomic DNA [5,7]. This phage replicates vegetatively and exclusively on *Brucella abortus* strains, resulting in the formation of clear plaques of approximately 5 mm [9]. Brucellaphage Pr was isolated in 2003 in Perote, Mexico. When replicating in *B. abortus* and *B. suis*, Pr forms clear plaques, and when replicating in *B. melitensis*, Pr produces small turbid plaques (Appendix A).

Most of the brucellaphages described so far infect only *B. abortus* [4]. There have been only two reports of phages infecting *B. melitensis*: phage Bk [10] and phage Iz [11]. However, because Bk was not isolated from natural samples, Pr was an attractive model that could be useful for studying the variation between brucellaphages that are isolated from different geographic locations and isolated at different times.

2.2. Characterization of phage DNA and virion proteins

We wanted to compare Tb and Pr at the molecular level to find the differences that might explain their host specificity. DNAs of Tb and Pr

* Corresponding author. Fax: +52 55 57473931.
E-mail addresses: vflorelo@gmail.com (V. Flores), ahide99@hotmail.com (A. López-Merino), menher@laguna.fmedic.unam.mx (G. Mendoza-Hernandez), gguarner@cinvestav.mx (G. Guarneros).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.01.001

were digested with *Hin*dIII, *Nde*I and *Mbo*I and resolved by agarose gel electrophoresis. We noted that Tb showed three more bands than Pr (Fig. 1A) when digested with *Hin*dIII. We also found differences in the migration of the fragments obtained with *Nde*I digestion, again with Tb exhibiting a larger genome. Finally, we found an extra band of ~1200 bp in the restriction pattern of Tb with *Mbo*I. These observations indicate that Tb and Pr are closely related, but not identical. The electrophoretic differences between Tb and Pr are due to two large insertions (see below) in the genome of Tb. No electrophoretic mobility shifts were observed after heating and slow cooling or fast cooling of the *Eco*RV restriction products, as observed with λ DNA. This suggests the absence of cohesive ends in the DNA molecules of brucellaphages (Appendix B). When we subjected the purified viral particles of Tb and Pr to SDS-PAGE, we noticed the absence of a band of approximately 69 kDa in the proteome of Tb (Fig. 1B). We also noticed an extra band of 14 kDa in Tb that could not be identified by mass spectrometry (data not shown). The 69 kDa band for Pr was identified as the product of ORF 21. The absence of this band for Tb is due to an insertion in ORF 21, which results in a larger protein. This larger protein produces a band that overlaps with that of the product of ORF 12, an observation verified by mass spectrometry.

2.3. Genome assembly and annotation

The complete genome sequences of both Tb and Pr were determined using the Applied Biosystems SOLiD technology. Approximately 3.5 million 50 nt sequence reads were assembled for each genome into single-contig sequences of 41,148 bp for Tb and 38,253 bp for Pr with a mean redundancy of 4000×. The restriction fragments predicted from the nucleotide sequences were in accordance with those observed experimentally, assuming a circular topology (Appendix C). Because no *cos* sites were observed in brucellaphage genomes (Appendix B), the topology could only be explained by permutations of the genomes. Manual sequencing using primer walking confirmed this observation (data not shown); however, we were not able to determine the extent of the permutations. We noticed that the sequence initiation in the SOLiD assembly of both genomes coincided with a putative replication origin. Mapping the reads to the assembled sequences and manually permuted sequences yielded very low coverage in the region corresponding to the putative replication origin. We do not know if this observation has a structural significance. Finally, the nucleotide numbers were assigned based on the primary SOLiD assembly of the genome of Pr. The genomes of the two phages share 99.87% nucleotide identity (see Materials and methods) apart from two large insertions in the genome of Tb: one of 447 bp and another of 2443 bp. The remaining differences consist of 15 single nucleotide polymorphisms (SNPs), 2 insertions/deletions (InDels) and 9 substitutions (Table 1). The average GC content of both phage genomes is 48%, which is significantly lower than that of *Brucella* genomes (57%). Fifty-seven putative protein-coding genes were found in each genome. ORFs 42 and 50 in the genome of Pr were interrupted by premature stop codons. Using a probabilistic model trained with known ribosome binding sites (consensus aggaggt) we detected ribosome binding sites (RBS) for fifty-three genes in each genome. ORFs 3 and 15 in both genomes, ORFs 14, 15 and 56 in Pr, and ORFs 15 and 57 in Tb showed no RBS signal. Four putative transcriptional promoters were identified on a structural basis using a neural network (see Materials and methods). They were located upstream of ORFs 1, 12 and 14 in both genomes (terminase complex, portal protein and structural module, respectively), and upstream of Pr ORF 57 and Tb ORF 58 (DNA primase/polymerase). The circular permutation of the genomes indicates that the promoters upstream of ORF 1 and ORF 57/58 represent divergent promoters of a transcriptional regulatory region. Four putative rho-independent transcriptional terminators (DNA segments with a stem-loop secondary structure followed by a thymine-rich region) were located in Tb, one of which is disrupted in Pr DNA due to a SNP. Based on the presence of a DnaA-like gene and nine DnaA boxes (*Brucella* DnaA box consensus: ttntccaca) [12], we detected a putative *oriC*-like origin of replication located between Pr ORFs 57 and 01 and Tb ORFs 58 and 01. Based on the molecular signatures and sequence similarity of each putative gene product to proteins in the databases, we identified two functional modules: morphogenesis/host lysis and DNA metabolism. Nine ORFs with no similarity to known sequences in the databases (ORFans) [13] were

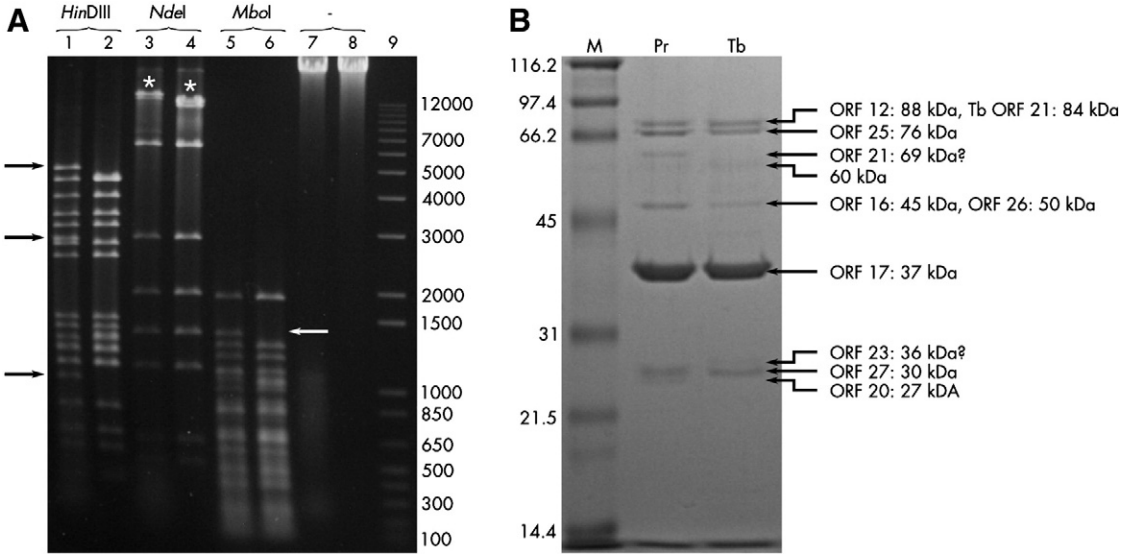

Fig. 1. Comparison of molecular patterns of brucellaphages Tb and Pr.(A) DNA from phages Tb (lanes 1, 3, 5 and 7) and Pr (lanes 2, 4, 6 and 8) were digested with *Hin*dIII (lanes 1 and 2), *Nde*I (lanes 3 and 4) and *Mbo*I (lanes 5 and 6) and subsequently subjected to gel electrophoresis for comparison of the resulting band patterns. The extra bands in the restriction patterns of Tb are indicated with arrows. The asterisks in the *Nde*I patterns show the differences in migration of the two large fragments from Tb with respect to those of Pr. Undigested DNAs (lanes 7 and 8) served as controls. Numbers on the right correspond to the sizes of the molecular marker (lane 9) in base pairs. (B) SDS-PAGE analysis of the viral particle proteins of Tb and Pr. Proteins were separated by acrylamide gel electrophoresis for 1.5 h at 180 volt and stained with Coomassie brilliant blue. The weights of the proteins were determined by comparison to the molecular weight marker (M). Proteins whose molecular weight did not match the theoretical weight are indicated with question marks. The proteins identified by mass spectrometry are shown on the right of the gel. ORF 21 in the proteome of Tb migrates the same distance as Tb ORF 12. The product of ORF 18 and the extra 14 kDa band of Tb are not shown because they were visible only if the proteins were separated for less than 1 h.

Table 1
Differences observed between the genome sequences of phages Tb and Pr.

Type of variation	Genome location (Tb position)	Observed sequence change		Blosum score
		Tb	Pr	
SNP	ORF 07, 08 (nt 3298)	N,P	Q,P	0,7
Substitution	Intergenic (ORFs 11–12, nt 6353–6357)	acccg	caaat	–
InDel	ORF 14 (nt 9188)	Start (M) at 9095	Start (L) at 9147	–
SNP	ORF 16 (nt 10,167)	F	V	−1
Substitution	Intergenic (ORFs 18–19, nt 12,382–12,388)	.atatac	aatat.g	–
Substitution	ORF 19 (nt 12,555–12,556)	R	Y	−2
Substitution	Intergenic (ORFs 19–20, nt 13,146–13,154)	..gaatagcac	atcgcta.ca.	–
Substitution	ORF 21 (nt 13,438–13,443)	HIC	RAR	−4
Substitution	ORF 21 (nt 13,449–13,450)	P	F	−4
SNP	ORF 21 (nt 14,455)	L	L	4
Major InDel	ORF 21 (nt 14,537–14,985)	Plus 149 aa	-	–
SNP	ORF 21 (nt 15,916)	A	V	−2
SNP	ORF 21 (nt 16,059)	P	T	1
SNP	ORF 23 (nt 16,678)	K	E	1
InDel	ORF 23 (nt 17,184–17,189)	LAVALT	LA..LT	–
SNP	ORF 27 (nt 21,792)	P	Q	−1
SNP	ORF 27 (nt 21,955)	N	K	0
SNP	ORF 27 (nt 22,364)	D	Y	−3
Substitution	ORF (nt 22,460–22,461)	S	D	0
SNP	ORF 27 (nt 22,466)	R	S	−1
SNP	ORF 27 (nt 22,470)	K	R	2
Major InDel	ORF 28 (nt 22,718–25,146)	Longer ORF 28 Plus 1 gene (Tb ORF 29)	–	–
Substitution	Intergenic (Tb ORFs 29–30, nt 25,151–24,152)	acgg.cga	acgggcga	–
SNP	Tb ORF 43 (nt 29,915)	E	Stop	–
Substitution	Tb ORF 51 (nt 34,533–34,537)	QIG	HR-Stop	−3
SNP	Tb ORF 58 (nt 39,320)	N	K	0
SNP	Genome terminus (nt 40,845)	g	t	–
SNP	Genome terminus (nt 41,105)	c	t	–

SNP: Single nucleotide polymorphism. InDel: Insertion or deletion. Capital letters indicate amino acid residues and small letters, nucleotide bases. "." = Missing residue. All coordinates and gene names are shown with respect to the Tb genome. The Blosum62 substitution matrix was used to assess the significance of the observed amino acid changes.

found, confirming that phages harbor a significant amount of uncharacterized sequences. A functional overview of both genomes is shown in Fig. 2 and summarized in Appendix D.

2.4. Morphogenesis

To identify the genes encoding the virion structural proteins, we subjected Tb and Pr CsCl-purified viral particles to SDS-PAGE. The resulting bands were trypsin-digested, and the obtained peptides were subjected to tandem mass spectrometry (see Materials and methods). Each phage displayed nine bands, from which we could identify the products of ORFs 12, 16, 17, 18, 20, 21, 23, 25, 26 and 27 in both genomes (Fig. 1 and Appendix D). We were not able to identify the gene encoding the 60 kDa band present in both phage proteomes either using OMSSA searches or from the predicted molecular weights of proteins encoded in the phage genomes. The predicted amino acid sequence of ORF 14 displayed conserved residues of the family Phage_head_fibr (Pfam acc. no. PF11133). This family includes head fibers of phages RSL1, infecting *Ralstonia* and phages Nf, PZA, B103 and φ29 infecting *Bacillus* [14]. Additionally, ORF 27 possesses the Phage Tail Collar domain (Pfam acc. no. PF07484). The size of the gene products determined by SDS-PAGE was close to that predicted from the DNA sequences, indicating that the structural proteins of both phages did not undergo significant post-translational processing. The putative structural proteins, ORFs 13, 14, 15, 19, 22 and 24, were not identified by mass spectrometry. This suggests that they were either absent in the viral particle or present, in low amount. Only ORFs 15 and 22 showed significant sequence similarity to known proteins. Sequence comparison revealed that most of the variations between Tb and Pr were observed in the virion structural proteins, whereas the rest of the genome showed little or no variability. Sequence pairwise comparison of the ORFs 21 in both phages showed a 149 amino acid insertion in ORF 21 of Tb. In addition, ORF 21 of Tb had six amino acid substitutions five of them could alter the protein function according to the score of matrix Blosum62 [15]. Also, we observed that Tb ORF 27 displayed six amino acid substitutions relative to that of Pr, three of them were predicted neutral for protein function [15]. We believe that ORFs 21 and 27 might be involved in host-specificity considering that one of the main differences at the phenotypic level between Tb and Pr is their host range. However experimental evidence is needed to test this hypothesis.

2.5. DNA packaging

DNA packaging in tailed bacteriophages is an important process in which DNA is accommodated into procapsids via ATP hydrolysis. Three main proteins are involved in DNA packaging: the terminase small and large subunits, and the ring-forming portal protein. Both terminase subunits remain in the cell, while the portal protein forms part of the mature virion [16]. Homology with proteins in the databases suggests that ORFs 6 and 7 of both brucellaphages encoded the terminase small and large subunits, respectively. The results from InterProScan confirmed our findings. On the other hand, the portal protein could not be identified by BLAST or InterProScan searches. However, the genes encoding the portal proteins being located immediately downstream of the terminase genes in phage genomes and the portal proteins being present in the mature viral particle [17,18], ORF 12 in brucellaphages appears to be the most-likely gene to encode the portal protein. In this case, it was demonstrated by mass spectrometry that the product of ORF 12 is present in the mature viral particles and that ORF 12 is downstream of the genes encoding the terminase complex. Although ORFs 9-11 are located immediately downstream of the terminase large subunit gene, comparison to a cryptic prophage in *Chelativorans* sp. BNC1 indicates that ORFs 9-11 correspond to an insertion. However it cannot be ruled out that the absence of these ORFs in the prophage resulted from a deletion. We considered it to be very unlikely that these genes encode the portal protein because they are very short compared to the mean

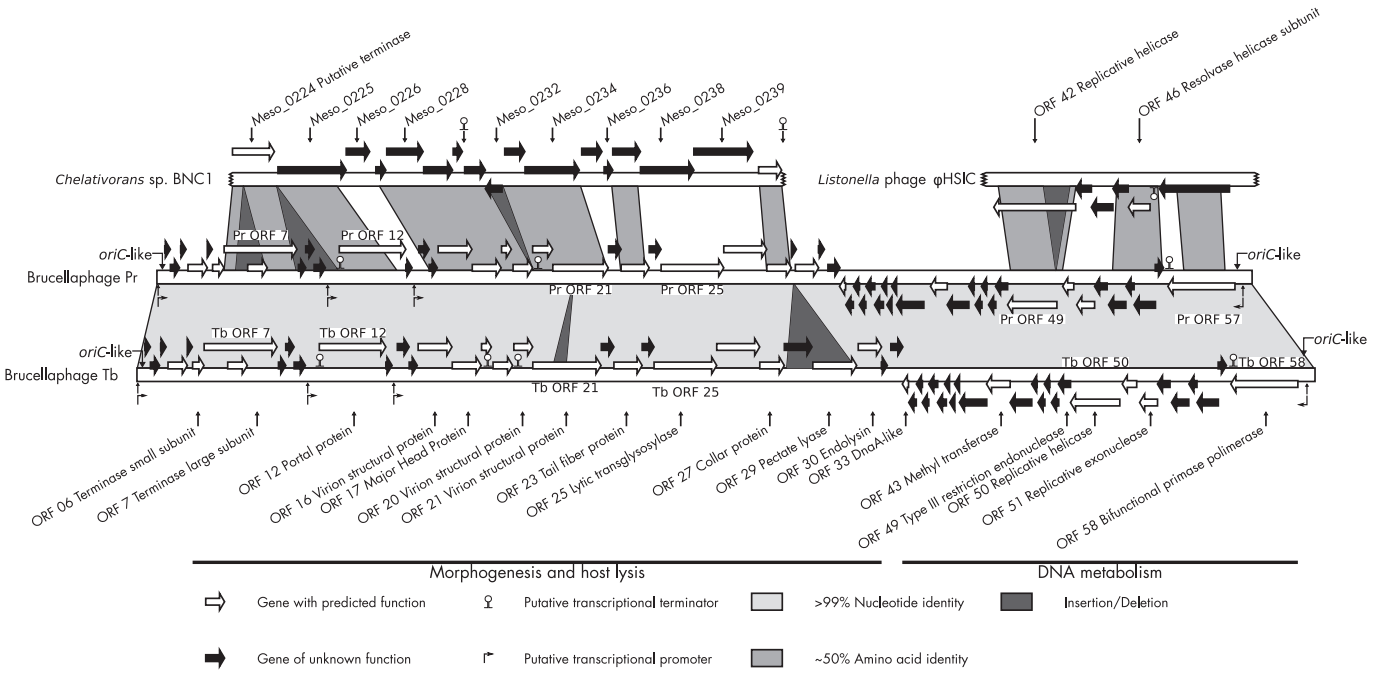

Fig. 2. Comparative genomics of brucellaphages, a cryptic prophage in *Chelativorans* BNC1 and *Listonella* phage φHSIC.The genomes are shown as white boxes. The circularly permuted sequences of brucellaphages genomes were opened at the first nucleotide in the SOLiD assembly of Pr. White arrows represent genes with an associated function, and black arrows genes of unknown function. Genes and genomes are drawn to scale. Promoters and terminators are marked with angled arrows and stem-loop signs, respectively. The putative *oriC*-like origin of replication (872 bp) starts at coordinates 37,655 and 40,550 in Pr and Tb, respectively, and ends at coordinate 273 in both genomes. The light gray areas connecting the brucellaphage genomes indicate nucleotide identity. Amino acid sequence similarity between brucellaphages and *Chelativorans* sp. BNC1 or *Listonella* phage φHSIC is shown as gray shaded areas between the corresponding genomes. Major InDels are depicted as dark gray triangles. Labels below and above the genomes indicate the putative function for the predicted genes. The lines below the genome of Tb show the modular architecture of brucellaphages genomes.

size of portal proteins (~480 aa). Moreover, they show little sequence similarity to proteins in the databases. The mechanism by which brucellaphages pack DNA inside empty heads was investigated indirectly using a neighbor-joining phylogeny of the amino acid sequence for the putative terminase large subunit [19]. The resulting tree (Fig. 3) revealed that brucellaphages may be related to phages that accommodate DNA in a headful manner, forming a previously undefined similarity cluster with *Lactococcus* phage TP901-1 and *Haemophilus* phage Aaφ23. In this context, the main characteristics of these phages are the circularly permuted genomes, the headful packaging of the DNA molecule and the absence of cohesive sites [19]. These findings provide information on the physiology of brucellaphages, which thus far has not been investigated.

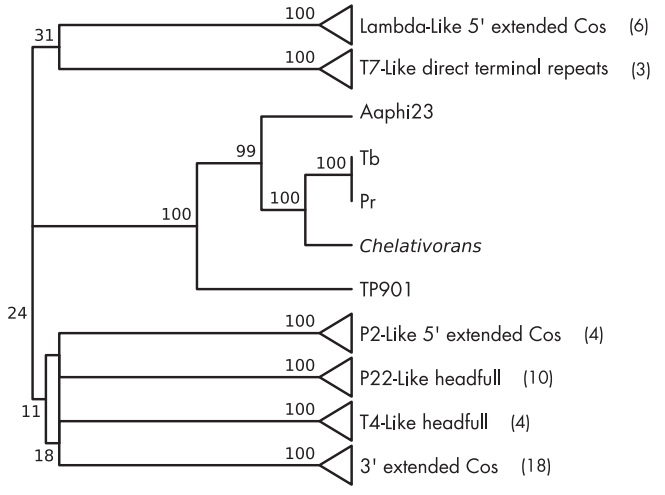

Fig. 3. Neighbor-joining tree of the terminase large subunit of brucellaphages.The tree is based on the alignment of the amino acid sequences of the terminase large subunit from a curated set of phages (12). Numbers on the nodes indicate the percent of replicas in which the corresponding groups were observed from a total of 1000 bootstraps. The main groups of DNA packaging strategies are shown at the right of the tree. The numbers in parentheses indicate the number of phages that form such groups. Brucellaphages, the cryptic prophage in *Chelativorans*, *Lactococcus* phage TP901 and *Haemophilus* phage Aaφ23 group into a previously undefined clade of headful-packing phages that is characterized by their circularly permuted genomes with no apparent *pac* site.

2.6. Host lysis

The products of Pr ORF 29 and Tb ORF 30 are possibly involved in host lysis, as they both possess a peptidoglycan binding domain and a DUF847 domain (Pfam accession numbers PF09374 and PF05838, respectively). These domains are typically observed in phage lytic proteins and bacterial peptidoglycan degrading enzymes [20]. The predicted amino acid sequence of these genes also displayed two putative transmembrane helices, a signature commonly found in phage holins [21]. Experiments are necessary to determine if the putative proteins are expressed and which activities are associated with them.

2.7. DNA modification and restriction

We found two genes encoding proteins that might be involved in DNA restriction and modification in both genomes. Pr ORF 42 and Tb ORF 43 encode a putative N4/N6 DNA methyl transferase (Mtase) and Pr ORF 48 and Tb ORF 49 encode a putative type III restriction nuclease. BLAST searches revealed that the Mtase shared amino acid sequence similarity to very diverse members of the NADP-Rossman protein clan, spanning distant lineages of bacteria and bacteriophages [22]. The same was observed for the type III restriction endonuclease that belongs to the PD-(D/E)XK protein clan [23]. These findings suggest that the genes involved in DNA modification and restriction belong to different lineages and that they might be related only in terms of their putative function. The presence of such genes is not

uncommon in tailed phage genomes [23]. There are five intervening ORFs between the putative nuclease and methylase genes whose function could not be inferred and their organization did not appear to reflect a conserved architecture.

2.8. Replication

The brucellaphages possess four genes that may be related to DNA replication. These are a putative DnaA-like protein (Pr ORF 32, Tb ORF 33), a putative replicative DEAD-Helicase (Pr ORF 49, Tb ORF 50), a PolB-related 3′-5′exonuclease (Pr ORF 51, Tb ORF 52) and a bifunctional DNA primase/polymerase (Pr ORF 57, Tb ORF 58).

DnaA proteins control the recruitment of the replicative machinery. This activity is mediated and regulated by its N-terminal portion, whereas DNA binding activity is mediated by its C-terminal region. The sequence of the DnaA-like protein of brucellaphages is considerably shorter than that of the canonical DnaA proteins due to their lack of the domains involved in protein binding and ATP hydrolysis; however, these domains have been demonstrated to be non-essential [24]. Although we do not have experimental evidence that supports the putative function of the DnaA-like protein, it might retain DNA binding activity because its C-terminal part is well conserved relative to that of *Mycoplama bovis* (YP_004683051; 60% amino acid similarity). Along with this putative protein, we identified a putative origin of replication to which DnaA might bind and promote replication events. This *oriC*-like locus spans 872 bp starting at 37,655 bp of Pr, or at 40,550 pb of Tb, and ends at 273 bp in the circularly permuted genomes (Fig. 2). The putative *oriC* contains nine DnaA boxes: five in the forward strand, and four in the reverse strand. DnaA boxes and DnaA protein are present in *Brucella* [12,25]. It would be interesting to investigate whether the two systems interact during phage DNA replication. There are no previous reports of the co-occurrence of DnaA-like proteins and *oriC*-like origins of replication in bacteriophage genomes. Nevertheless, ORFs encoding DnaA-like proteins have been found in phages infecting *Staphylococcus aureus*. The presence of such elements in brucellaphages suggests that these viruses might undergo DNA replication in a fashion resembling that of bacterial chromosomes. DNA replication is usually achieved by the activities of three enzymes: a DNA helicase that unwinds the double strand molecule, a DNA primase that synthesizes RNA primers to enable DNA polymerization and a DNA polymerase that elongates the primed DNA molecule [26]. Brucellaphages DNA replication might be accomplished by a DEAD helicase and a bifunctional enzyme with DNA primase and polymerase activities [26]. This type of enzymes (represented by the *Sulfolobus islandicus* plasmid replicase NP_044372) is commonly found in phage genomes, and they appear to have a relationship to archaeal and eukaryotic plasmid replicases [27]. We also found a putative PolB-related 3′-5′ exonuclease that might serve as an accessory protein in charge of the proofreading activity of the DNA primase/polymerase. Between the exonuclease and polymerase ORFs, there were five genes whose functions could not be identified. However, this organization seems to be a conserved feature because the replication module of brucellaphages resembles that of *Listonella* phage φHSIC (see below).

2.9. Genome architecture

The genomes of brucellaphages are organized in two arms: morphogenesis/host lysis and DNA metabolism. All of the genes found in the module of morphogenesis/host lysis are located in the forward strand, whereas those in the DNA metabolism are on the reverse strand. Both genomes possess 59% of small clustered genes whose functions are unknown and that show little similarity to other sequences. Small genes in phages have been noticed and discussed before [28].

2.10. Genome comparison

We conducted a pairwise comparison between Tb and Pr genomes, and as expected from DNA restriction profiles, we observed a high degree of shared identity between both genomes (99.87%, excluding the two large insertions in Tb). This strikingly high level of genotypic homology, despite the significant differences in geographic and temporal conditions at which the phages were isolated, suggests that brucellaphages are among the most conserved phages. Traditionally it was believed that lytic phages isolated at different places and at different times were not likely to show significant nucleotide similarity in spite of infecting the same host [29]; however, recent evidence proves that there are indeed phages isolated at distant regions with highly similar genomes ([30] and references therein). Based on our observations, we propose that brucellaphages are so well adapted to their hosts that most of their genes are conserved and that only those genes possibly involved in host-specificity show some variability (see above). By comparison of the genome sequences of brucellaphages to sequences in the databases, we detected strong similarity at both the amino acid level (~50%) and gene order, to a possible cryptic prophage in the genome of *Chelativorans* sp. BNC1. This suggests an evolutionary relationship in which brucellaphages and the cryptic prophage might have shared a common ancestor. This is further supported by evidence that *Brucella* and *Chelativorans* are closely related; they both belong to the order *Rhizobiales* [31]. Analyses of *Brucella* genome sequences have revealed the presence of cryptic prophages [32]; however, these prophages show neither sequence similarity nor conserved architecture relative to the lytic brucellaphages or the cryptic prophage in *Chelativorans*. Additionally, the sequence similarity between brucellaphages and the cryptic prophage occurs at the amino acid level, but not at the nucleotide level, indicating an ancient divergence. Finally, the module of DNA replication of brucellaphages, excluding the putative DnaA ORF, showed ~50% amino acid sequence similarity and conserved gene organization to that of *Listonella* phage φHSIC. Because *Listonella* is a marine *Gammaproteobacteria* and *Brucella* and *Chelativorans* are terrestrial *Alphaproteobacteria*, it is difficult to establish a coherent hypothesis for the relationship of their phages.

2.11. Gene-loss and -acquisition

Comparison of the amino acid sequences of the putative terminase large subunit in brucellaphages (ORF 7) and the gene-product of Meso_0224 of *Chelativorans* revealed an in-frame insertion that corresponds to a putative intein. All inteins have conserved amino acid residues that can be used to identify the position at which they are inserted in the host protein [33]. We found the C-terminal of the intein by the conserved motif Asn-Cys at position 506 in ORF 7. Sequence similarity to Meso_0224 (67%) indicated that the N-terminal was the serine at position 185. The intein sequence also appears to have an insertion, but we did not find intein-specific conserved residues in the second insertion. We also found an ORF encoding a homing endonuclease, which was out of frame relative to the second insertion in the terminase gene. Hence, we hypothesize that this second insertion corresponds to a prokaryotic intron [34]. The occurrence of introns in phage large terminase genes has been noticed [28]. Inteins and introns are selfish elements that are difficult to purge from prokaryotic genomes [33]. The presence of these intervening sequences in the genome of brucellaphages, but not in the cryptic prophage, indicate that their putative common ancestor presumably did not have the intein or the intron. Another possible gene acquisition event might have taken place between ORFs 7 and 12 because ORFs 9-11 did not show sequence similarity to *Chelativorans* sp. In contrast, Brucellaphages differ from the cryptic phage in *Chelativorans* sp. BNC in that they lack the hypothetical protein encoded in Meso_0232. The amino acid sequence of this putative

protein displays a helix-turn-helix DNA binding motif, suggesting a function as a phage repressor [35].

2.12. Conclusions

We determined the genome sequences of two distantly isolated brucellaphages that differ in their host range. Genome analysis revealed that brucellaphages are unique in their strikingly high degree of conservation and their putative mechanism of DNA replication initiation. Genome comparison provided information concerning the evolutionary relationships between brucellaphages and a putative cryptic prophage in *Chelativorans* sp. BNC1 and the pseudotemperate phage, φHSIC of *Listonella*. Genome organization revealed that brucellaphages have an architecture of genes involved in DNA packaging – morphogenesis – host lysis in the forward strand and DNA metabolism in the reverse strand. This architecture is conserved among some phages with bifunctional DNA primases/polymerases; however, we were not able to detect any evolutionary pattern. Functional analysis revealed a high percentage of putative proteins of unknown functions. The genes encoding these putative proteins were scattered along the genome, and many of them appear to have originated from genomic insertion. Finally, we propose that the features of brucellaphages are the result of a high level of adaptation to their hosts, with only a fraction of their genes showing minor variations.

3. Materials and methods

3.1. Bacterial and brucellaphage strains

All the bacteria utilized here were purchased from the Central Veterinary Laboratory, Weybridge England: *Brucella abortus* bv. 1 strain 544, *Brucella melitensis* bv. 1 strain 16 M and *Brucella suis* bv. 1 strain 1330. Bacteria were grown in *Brucella* broth (Oxoid Limited, Hampshire, UK), supplemented with 0.5% fetal bovine serum (Sigma-Aldrich, St. Louis, MO, USA), 0.7% yeast extract and 0.5% casamino acids (BD-Difco, Oxford, UK), and agar was added to a final concentration of 2%. Phage Tb was purchased from the Gamaleya Scientific Research Institute of Epidemiology and Microbiology, Moscow (www.gamaleya.ru). Using standard procedures [4], Phage Pr was isolated in 2003 by Nidia Guadalupe Aréchiga Ceballos during her master's degree experimental work at the Laboratorio de Microbiología General of the Escuela Nacional de Ciencias Biológicas (ENCB IPN). Due to its broad host-range, Pr was selected from the ENCB IPN collection of brucellaphages isolated in México.

3.2. Phage specificity

We propagated single plaques of Pr and Tb in *B. abortus* for three consecutive rounds in 5 mL of tryptic soy broth. After centrifugation for 10 minutes at 10,000×*g*, the supernatants from the last lysates were filtered through a 0.22 μm filter and diluted up to 10^{-5}. To determine Pr specificity, we spotted 5 μL of each dilution onto lawns of *B. abortus*, *B. melitensis* and *B. suis*. Pr was considered infective if plaques formed on the tested strain.

3.3. Phage purification

Brucellaphage Tb was propagated in *B. abortus* 544, whereas Pr was propagated in *B. melitensis* 16 M. Large scale phage propagation was carried out by inoculating 3 L of TSB-YE with 50 mL of a bacterial suspension containing 9×10^8 CFU/mL and 5 mL of brucellaphage suspension at approximately10^5 PFU/mL. After 72 h of incubation at 37 °C with constant shaking, DNAse I and RNAse A were added (625 μg each), and cellular debris was removed by centrifugation for 10 min at 10,000×*g*. The supernatant was filtered using Steri-Cup® vacuum filtering systems (Millipore). Polyethylene glycol 8000 (Invitrogen Corp., CA, USA) and NaCl (Mallinckrodt Baker, NJ, USA) were added to the filtered lysate at a final concentration of 20% and 1 M, respectively. The mixtures were left overnight at 4 °C in centrifuge bottles. Phage particles were sedimented by centrifugation at 15,000×*g* for 1 h at 4 °C. The supernatant for each was discarded, and the pelleted phages were resuspended in 2 mL of Buffer SM (NaCl 0.1 M, $MgSO_4$ 0.01 M, Tris-Cl 0.05 M and gelatin 0.01%). PEG was removed by chloroform extraction. The resulting phage suspensions were placed on top of a CsCl step density gradient (3.5 mL of ρ = 1.7 g/mL; 2.5 mL of ρ = 1.5 g/mL; 2.5 mL of ρ = 1.3 g/mL; and 2.5 mL of ρ = 1.08 g/mL) and centrifuged for 2.5 h at 86,000×*g* in a Beckman SW-40 Ti rotor. Phage bands were aspired and dialyzed against 2 L of Tris-Cl 50 mM, NaCl 10 mM and $MgCl_2$ 10 mM overnight.

3.4. DNA isolation, RFLP, sequencing and assembly

DNA was obtained by phenol-chloroform extraction from the purified phage suspensions as described previously [36]. Enzymatic restrictions were performed by mixing 1 μg of purified DNA with 1 unit of *Eco*RV, *Hin*dIII, *Mbo*I or *Nde*I (New England Biolabs, Ipswich, MA, USA). Restriction fragments were resolved on a 1.5% agarose gel and stained with ethidium bromide. To investigate the presence of cohesive ends in the brucellaphages genomes, EcoRV restriction fragments were heated for 10 min at 75 °C and then quickly cooled. Unheated and slowly cooled reactions were used as controls. The gel migration patterns were compared to those obtained with DNA of phage λ under the same conditions. High-throughput SOLiD sequencing was carried out at the National Laboratory of Genomics for Biodiversity (LANGEBIO-CINVESTAV, Irapuato, Mexico). Genomic libraries were constructed from 10 μg of purified DNA according to the Barcode protocol of Applied Biosystems (Carlsbad, CA, USA). Emulsion PCR reactions were prepared manually and analyzed using a SOLiD 3.5 System. Sequence reads were preprocessed using the Applied Biosystems de novo assembly accessories and assembled using Velvet v1.1 [37]. Refinement of the assembly was performed by inspection, and the final sequences were completed by Sanger sequencing.

3.5. Bioinformatics and genome annotation

Open reading frames were predicted with heuristic Hidden Markov Models using GeneMark v1.1 [38]. Ribosome binding sites were verified with rbs_finder.pl [39] for correct location of ORF positions. A search for transfer RNAs was performed using tRNAscan-SE v1.23 [40]. Rho-independent transcriptional terminators were detected using transterm-hp v2.07 [41]. Non-coding sequences were submitted to the Berkeley Drosophila Genome Project website for promoter prediction [42] (http://www.fruitfly.org/seq_tools/promoter.html). Origins of replication were searched using the Orifinder web server [43] (http://tubic.tju.edu.cn/Ori-Finder/). Homologs for the predicted genes were identified via BLAST searches [44] against the non-redundant protein database on the NCBI server. Conserved protein domains and protein families were searched with InterProScan [45] and NCBI-CDD [46]. Functional genome annotation was achieved by integrating BLAST, InterPro and CDD data with the use of the Artemis Annotation Tool [47]. General features of the nucleotide and amino acid sequences were determined using the EMBOSS package [48].

3.6. DNA packaging

The DNA packaging mechanism of brucellaphage relative to other known phages was investigated by constructing a neighbor-joining phylogeny from a set of curated sequences based on the large subunit of their terminase [19]. Alignment of the amino acid sequences was performed with MUSCLE v3.8 [49]. Refinement of the alignments

and neighbor-joining reconstruction were performed with Seaview v4.2 [50].

3.7. Genome comparison

The genomes of both Tb and Pr were compared with the sequence of a putative cryptic prophage in the genome of *Chelativorans* sp. BNC1 (accession: NC_008254) and with the terminal region of the genome of *Listonella* phage φHSIC [51] (GenBank acc. no. NC_006953) using MUMmer [52] and tBLASTx [44]. Finally, comparative maps were constructed using in-house scripts and the Artemis Comparison Tool [47].

3.8. Protein analysis and proteomics

Purified bacteriophages were mixed with Laemmli loading buffer and boiled for 5 min. The mixture was loaded onto a 10% SDS-glycerol-acrylamide gel. Proteins were resolved for 1.5 h at 180 volt and stained with Coomassie brilliant blue [34], another gel was run for 1 h at 180 volt to identify low molecular weight proteins. The molecular weight of the observed proteins was estimated by comparison with a pre-stained SDS-PAGE broad range protein standard (BioRad Hercules, CA, USA. Cat. No. 161-0317). The protein bands were carefully excised from the Coomassie-stained SDS gel and destained with a mixture of 50% methanol and 5% acetic acid for 12 h. The destained slices were washed with deionized water, soaked for 10 min in 100 mM ammonium bicarbonate, dehydrated with 100% acetonitrile and vacuum-dried. Proteins were reduced with 10 mM DTT and S-alkylated cysteine with 100 mM iodoacetamide in 100 mM ammonium bicarbonate. In-gel digestion was performed by adding 600 ng of mass spectrometry-grade trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate followed by overnight incubation at room temperature. Peptides were extracted twice with 50% acetonitrile and 5% formic acid for 30 min. The extracts were vacuum-dried and resuspended in 20 μL of 0.1% formic acid. Analysis of tryptic peptides was carried out using an integrated nano-LC-ESI-MS/MS system. The raw data files were processed and converted to peak lists in pkl format. Protein identification was achieved with OMSSA to compare the obtained spectra to predicted amino acid sequences for Tb and Pr genomes [53].

3.9. Accession numbers

The nucleotide sequences and annotations of the brucellaphages genomes were deposited in GenBank with accession numbers JN939332 and JN939331 for Pr and Tb, respectively.

Acknowledgments

V. Flores was a recipient of a Master degree fellowship from the Consejo Nacional de Ciencia y Tecnología (CONACYT) and he is currently a doctoral fellowship recipient from CONACYT. This work was in part supported by CONACyT and the Instituto de Ciencia y Tecnología del Distrito Federal (ICyTDF). We gratefully thank Omar Sepúlveda for the purification of brucellaphage Tb; and Guadalupe Aguilar for the Sanger sequencing services. We also thank the anonymous reviewer whose comments and suggestions were taken into account in the final version of the manuscript.

Appendix A. Host range of Pr and Tb

Phage lysate was spotted onto lawns of each *Brucella* species at the dilutions indicated in the left side of the diagram. Phages were considered infective by the presence of lysis zones. Note the turbid lytic spots of Pr on *B. melitensis* lawn.

Appendix B. Genome ends in brucellaphages

DNAs from phages Pr (lanes 1–3), Tb (lanes 4–6), and λ (lanes 7–9) were digested with *Eco*RV, heated for 10 min at 75 °C and quickly cooled to 4 °C (lanes 2, 5 and 8) or slowly cooled (lanes 3, 6 and 9) before gel electrophoresis. Unheated restriction fragments were used as controls (lanes 1, 4 and 7). The numbers on the right correspond to the size of the molecular marker in base pairs. The 3000 bp fragment in the λ restriction pattern (black arrow) dissociates into two fragments (white arrows), indicating the presence of cohesive ends in the λ genome. No changes were detected in the restriction patterns of brucellaphages. Thus, we conclude that they do not possess cohesive ends in their DNA.

Appendix C. Correlation between the observed restriction fragments and those predicted assuming a linear or a circular topology of the DNA molecule

"−" = Restriction fragments that were not observed in gel electrophoresis. "+" = Restriction fragments observed in gel electrophoresis. [a] Restriction fragments that could not be observed due to their small size. [b] Restriction fragments predicted from the linear topology of the molecule, absent in gel electrophoresis.

Appendix D. Functional overview of brucellaphages genomes

Proteins denoted as "Structural protein" were identified by mass spectrometry. Genes denoted as ORFans showed no similarity to known sequences in databases.

Appendix E. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2012.01.001.

References

[1] J. Godfroid, A. Cloeckaert, J.-P. Liautard, S. Kohler, D. Fretin, K. Walravens, B. Garin-Bastuji, J.-J. Letessonf, From the discovery of the Malta fever's agent to the discovery of a marine mammal reservoir, brucellosis has continuously been a re-emerging zoonosis, Vet. Res. 36 (2005) 313–326.
[2] M.P. Franco, M. Mulder, R.H. Gilman, H.L. Smits, Human brucellosis, Lancet Infect. Dis. 7 (2007) 775–786.
[3] A.M. Whatmore, Current understanding of the genetic diversity of *Brucella*, an expanding genus of zoonotic pathogens, Infect. Genet. Evol. 9 (2009) 1168–1184.
[4] M.J. Corbel, E.L. Thomas, The Brucella-Phages: Their Properties, Characterisation and Applications, Ministry of Agriculture, Fisheries and Food, Weybridge, Surrey, England, 1980.
[5] J.G. Calderone, M.J. Pickett, Characterization of brucellaphages, J. Gen. Microbiol. 39 (1965) 1–10.
[6] J.A. Morris, M.J. Corbel, J.I. Phillip, Characterization of three phages lytic for *Brucella* species, J. Gen. Virol. 20 (1973) 63–73.
[7] C.E. Rigby, M.L. Cerqueira-Campos, H.a. Kelly, O.P. Surujballi, Properties and partial genetic characterization of Nepean phage and other lytic phages of *Brucella* species, Can. J. Vet. Res. 53 (1989) 319–325.
[8] M. Segondy, Common physical map of four *Brucella* bacteriophage genomes, FEMS Microbiol. Lett. 56 (1988) 177–181.
[9] C.R. McDuff, L.M. Jones, J.B. Wilson, Characteristics of brucellaphage, J. Bacteriol. 83 (1962) 324–329.
[10] J.T. Douglas, S.S. Elberg, Isolation of *Brucella melitensis* phage of broad biotype and species specificity, Infect. Immun. 14 (1976) 306–308.
[11] M.J. Corbel, F. Tolari, V.K. Yadava, Characterisation of a new phage lytic for both smooth and non-smooth *Brucella* species, Res. Vet. Sci. 44 (1988) 45–49.
[12] P. Mackiewicz, J. Zakrzewska-Czerwinska, A. Zawilak, M.R. Dudek, S. Cebrat, Where does bacterial replication start? Rules for predicting the oriC region, Nucleic Acids Res. 32 (2004) 3781–3791.
[13] D. Fischer, D. Eisenberg, Finding families for genomic ORFans, Bioinformatics 15 (1999) 759–762.
[14] Y. Tao, N.H. Olson, W. Xu, D.L. Anderson, M.G. Rossmann, T.S. Baker, Assembly of a tailed bacterial virus and its genome release studied in three dimensions, Cell 95 (1998) 431–437.
[15] S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 10915–10919.
[16] V.B. Rao, M. Feiss, The bacteriophage DNA packaging motor, Annu. Rev. Genet. 42 (2008) 647–681.

[17] J.M. Valpuesta, J.L. Carrascosa, Structure of viral connectors and their function in bacteriophage assembly and DNA packaging, Q. Rev. Biophys. 27 (1994) 107–155.
[18] A. Zachary, L.W. Black, Isolation and characterization of a portal protein-DNA complex from dsDNA bacteriophage, Intervirology 33 (1992) 6–16.
[19] S.R. Casjens, E.B. Gilcrease, D.A. Winn-stapley, P. Schicklmaier, H. Schmieger, M.L. Pedulla, M.E. Ford, J.M. Houtz, G.F. Hatfull, R.W. Hendrix, The generalized transducing salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy, J. Bacteriol. 187 (2005) 1091–1104.
[20] J. Pei, N.V. Grishin, COG3926 and COG5526: a tale of two new lysozyme-like protein families, Protein Sci. 14 (2005) 2574–2581.
[21] I.N. Wang, D.L. Smith, R. Young, Holins: the protein clocks of bacteriophage infections, Annu. Rev. Microbiol. 54 (2000) 799–825.
[22] M. Bashton, C. Chothia, The geometry of domain combination in proteins, J. Mol. Biol. 315 (2002) 927–939.
[23] L.N. Kinch, K. Ginalski, L. Rychlewski, N.V. Grishin, Identification of novel restriction endonuclease-like fold families among hypothetical proteins, Nucleic Acids Res. 33 (2005) 3598–3605.
[24] W. Messer, The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication, FEMS Microbiol. Rev. 26 (2002) 355–374.
[25] S.M. Halling, B.D. Peterson-Burch, B.J. Bricker, R.L. Zuerner, Z. Qing, L.L. Li, V. Kapur, D.P. Alt, S.C. Olsen, Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*, J. Bacteriol. 187 (2005) 2715–2726.
[26] A. Kornberg, T. Baker, DNA Replication, W.H. Freeman, New York, USA, 1992.
[27] G. Lipps, A.O. Weinzierl, G. von Scheven, C. Buchen, P. Cramer, Structure of a bifunctional DNA primase-polymerase, Nat. Struct. Mol. Biol. 11 (2004) 157–162.
[28] C.R. Stewart, S.R. Casjens, S.G. Cresawn, J.M. Houtz, A.L. Smith, M.E. Ford, C.L. Peebles, G.F. Hatfull, R.W. Hendrix, W.M. Huang, M.L. Pedulla, The genome of *Bacillus subtilis* bacteriophage SPO1, J. Mol. Biol. 388 (2009) 48–70.
[29] H. Brüssow, R.W. Hendrix, Phage genomics: small is beautiful, Cell 108 (2002) 13–16.
[30] A. Carrias, T.J. Welch, G.C. Waldbieser, D.A. Mead, J.S. Terhune, M.R. Liles, Comparative genomic analysis of bacteriophages specific to the channel catfish pathogen *Edwardsiella ictaluri*, Virol. J. 8 (2011).
[31] F.M. Carvalho, R.C. Souza, F.G. Barcellos, M. Hungria, A.T.R. Vasconcelos, Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order *Rhizobiales*, BMC Microbiol. 10 (2010).
[32] H.C. Scholz, Z. Hubalek, I. Sedlácek, G. Vergnaud, H. Tomaso, S. Al Dahouk, F. Melzer, P. Kämpfer, H. Neubauer, A. Cloeckaert, M. Maquart, M.S. Zygmunt, A.M. Whatmore, E. Falsen, P. Bahn, C. Göllner, M. Pfeffer, B. Huber, H.-J. Busse, K. Nöckler, *Brucella microti* sp. nov., isolated from the common vole *Microtus arvalis*, Int. J. Syst. Evol. Microbiol. 58 (2008) 375–382.
[33] R. Raghavan, M.F. Minnick, Group I introns and inteins: disparate origins but convergent parasitic strategies, J. Bacteriol. 191 (2009) 6193–6202.
[34] B. Dassa, N. London, B.L. Stoddard, O. Schueler-Furman, S. Pietrokovski, Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family, Nucleic Acids Res. 37 (2009) 2560–2573.
[35] L. Aravind, V. Anantharaman, S. Balaji, M.M. Babu, L.M. Iyer, The many faces of the helix-turn-helix domain: transcription regulation and beyond, FEMS Microbiol. Rev. 29 (2005) 231–262.
[36] J. Sambrook, D. Russell, Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, New York, USA, 2001.
[37] D.R. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829.
[38] M. Borodovsky, R. Mills, J. Besemer, A. Lomsadze, Prokaryotic Gene Prediction Using GeneMark and GeneMark.hmm, in, 2003, , 2003.
[39] B.E. Suzek, M.D. Ermolaeva, M. Schreiber, S.L. Salzberg, A probabilistic method for identifying start codons in bacterial genomes, Bioinformatics 17 (2001) 1123–1130.
[40] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
[41] C.L. Kingsford, K. Ayanbule, S.L. Salzberg, Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake, Genome Biol. 8 (2007) R22–R.
[42] M.G. Reese, Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome, Comput. Chem. 26 (2001) 51–56.
[43] F. Gao, C.T. Zhang, Ori-Finder: a web-based system for finding *oriC*s in unannotated bacterial genomes, BMC Bioinformatics 9:79 (2008).
[44] S.F. Altschul, T.L. Madden, a.a. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[45] S. Hunter, R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A.F. Quinn, J.D. Selengut, C.J.A. Sigrist, M. Thimma, P.D. Thomas, F. Valentin, D. Wilson, C.H. Wu, C. Yeats, InterPro: the integrative protein signature database, Nucleic Acids Res. 37 (2009) D211–D215.
[46] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.a. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a conserved domain database for the functional annotation of proteins, Nucleic Acids Res. 39 (2011) D225–D229.
[47] T. Carver, M. Berriman, A. Tivey, C. Patel, U. Böhme, B.G. Barrell, J. Parkhill, M.-A. Rajandream, Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database, Bioinformatics 24 (2008) 2672–2676.
[48] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, Trends Genet. 16 (2000) 276–277.
[49] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinforma. 5 (2004).
[50] M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, Mol. Biol. Evol. 27 (2010) 221–224.
[51] J.H. Paul, S.J. Williamson, A. Long, R.N. Authement, D. John, A.M. Segall, F.L. Rohwer, M. Androlewicz, S. Patterson, Complete genome sequence of φHSIC, a pseudotemperate marine phage of *Listonella pelagia*, Appl. Environ. Microbiol. 71 (2005) 3311–3320.
[52] S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S.L. Salzberg, Versatile and open software for comparing large genomes, Genome Biol. 5 (2004) R12–R.
[53] L.Y. Geer, S.P. Markey, J.a. Kowalak, L. Wagner, M. Xu, D.M. Maynard, X. Yang, W. Shi, S.H. Bryant, Open mass spectrometry search algorithm, Proteome Res. 3 (2004) 958–964.